(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

(11) **N° de publication :** **2 768 544**
(à n'utiliser que pour les commandes de reproduction)

(21) **N° d'enregistrement national :** **97 11640**

(51) Int Cl[6] : **G 10 L 3/00**, G 10 L 7/06

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt :** 18.09.97.

(30) **Priorité :**

(43) **Date de mise à la disposition du public de la demande :** 19.03.99 Bulletin 99/11.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Se reporter à la fin du présent fascicule*

(60) **Références à d'autres documents nationaux apparentés :**

(71) **Demandeur(s) :** *MATRA COMMUNICATION SOCIETE PAR ACTIONS SIMPLIFIEE* — FR.

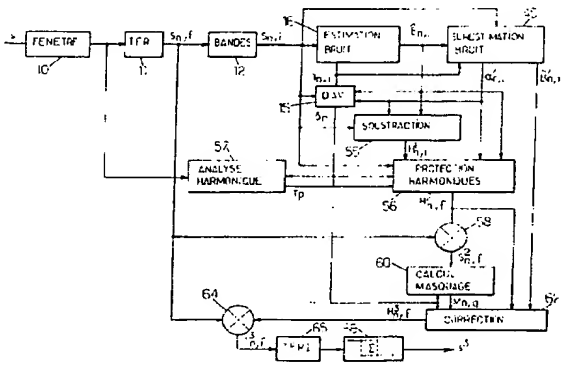
(72) **Inventeur(s) :** LOCKWOOD PHILIP et LUBIARZ STEPHANE.

(73) **Titulaire(s) :**

(74) **Mandataire(s) :** CABINET PLASSERAUD.

(54) PROCEDE DE DETECTION D'ACTIVITE VOCALE.

(57) Le signal de parole numérique (s) traité par trames successives est soumis à un débruitage en tenant compte d'estimations du bruit compris dans le signal, mises à jour pour chaque trame d'une manière dépendante d'au moins un degré d'activité vocale ($\gamma_{n,i}$). On procède à un débruitage a priori du signal de parole de chaque trame sur la base d'estimations du bruit obtenues lors du traitement d'au moins une trame précédente, et on analyse les variations d'énergie du signal débruité a priori pour détecter le degré d'activité vocale de ladite trame.



FR 2 768 544 - A1

PROCEDE DE DETECTION D'ACTIVITE VOCALE

La présente invention concerne les techniques numériques de traitement de signaux de parole. Elle concerne plus particulièrement les techniques faisant appel à une détection d'activité vocale afin d'effectuer des traitements différenciés selon que le signal supporte ou non une activité vocale.

Les techniques numériques en question relèvent de domaines variés : codage de la parole pour la transmission ou le stockage, reconnaissance de la parole, diminution du bruit, annulation d'écho...

Les méthodes de détection d'activité vocale ont pour principale difficulté la distinction entre l'activité vocale et le bruit qui l'accompagne. Le recours à une technique de débruitage classique ne permet pas de traiter cette difficulté, puisque ces techniques font elles-mêmes appel à des estimations du bruit qui dépendent du degré d'activité vocale du signal.

Un but principal de la présente invention est d'améliorer la robustesse au bruit des méthodes de détection d'activité vocale.

L'invention propose ainsi un procédé de détection d'activité vocale dans un signal de parole numérique traité par trames successives, dans lequel on soumet le signal de parole à un débruitage en tenant compte d'estimations du bruit compris dans le signal, mises à jour pour chaque trame d'une manière dépendante d'au moins un degré d'activité vocale déterminé pour ladite trame. Selonl'invention, on procède à un débruitage a priori du signal de parole de chaque trame sur la base d'estimations du bruit obtenues lors du traitement d'au moins une trame précédente, et on analyse les variations d'énergie du signal débruité a priori pour détecter le degré d'activité

vocale de ladite trame.

Le fait de procéder à la détection d'activité vocale (selon une méthode qui peut généralement être toute méthode connue) sur la base d'un signal débruité a priori améliore sensiblement les performances de cette détection lorsque le bruit environnant est relativement important.

Dans la suite de la présente description, on illustrera le procédé de détection d'activité vocale selon l'invention dans un système de débruitage d'un signal de parole. On comprendra que ce procédé peut trouver des applications dans de nombreux autres types de traitement numérique de la parole dans lesquels on souhaite disposer d'une information sur le degré d'activité vocale du signal traité : codage, reconnaissance, annulation d'écho...

D'autres particularités et avantages de la présente invention apparaîtront dans la description ci-après d'exemples de réalisation non limitatifs, en référence aux dessins annexés, dans lesquels :

- la figure 1 est un schéma synoptique d'un système de débruitage mettant en œuvre la présente invention ;
- les figures 2 et 3 sont des organigrammes de procédures utilisées par un détecteur d'activité vocale du système de la figure 1 ;
- la figure 4 est un diagramme représentant les états d'un automate de détection d'activité vocale ;
- la figure 5 est un graphique illustrant les variations d'un degré d'activité vocale ;
- la figure 6 est un schéma synoptique d'un module de surestimation du bruit du système de la figure 1 ;
- la figure 7 est un graphique illustrant le calcul d'une courbe de masquage ; et
- la figure 8 est un graphique illustrant l'exploitation des courbes de masquage dans le système de

la figure 1.

Le système de débruitage représenté sur la figure 1 traite un signal numérique de parole s. Un module de fenêtrage 10 met ce signal s sous forme de fenêtres ou trames successives, constituées chacune d'un nombre N d'échantillons de signal numérique. De façon classique, ces trames peuvent présenter des recouvrements mutuels. Dans la suite de la présente description, on considérera, sans que ceci soit limitatif, que les trames sont constituées de N=256 échantillons à une fréquence d'échantillonnage F_e de 8 kHz, avec une pondération de Hamming dans chaque fenêtre, et des recouvrements de 50% entre fenêtres consécutives.

La trame de signal est transformée dans le domaine fréquentiel par un module 11 appliquant un algorithme classique de transformée de Fourier rapide (TFR) pour calculer le module du spectre du signal. Le module 11 délivre alors un ensemble de N=256 composantes fréquentielles du signal de parole, notées $S_{n,f}$, où n désigne le numéro de la trame courante, et f une fréquence du spectre discret. Du fait des propriétés des signaux numériques dans le domaine fréquentiel, seuls les N/2=128 premiers échantillons sont utilisés.

Pour calculer les estimations du bruit contenu dans le signal s, on n'utilise pas la résolution fréquentielle disponible en sortie de la transformée de Fourier rapide, mais une résolution plus faible, déterminée par un nombre I de bandes de fréquences couvrant la bande $[0,F_e/2]$ du signal. Chaque bande i ($1 \leq i \leq I$) s'étend entre une fréquence inférieure f(i-1) et une fréquence supérieure f(i), avec f(0)=0, et $f(I)=F_e/2$. Ce découpage en bandes de fréquences peut être uniforme ($f(i)-f(i-1)=F_e/2I$). Il peut également être non uniforme

(par exemple selon une échelle de barks). Un module 12 calcule les moyennes respectives des composantes spectrales $S_{n,f}$ du signal de parole par bandes, par exemple par une pondération uniforme telle que :

$$S_{n,i} = \frac{1}{f(i) - f(i-1)} \sum_{f \in [f(i-1), f(i)[} S_{n,f} \qquad (1)$$

Ce moyennage diminue les fluctuations entre les bandes en moyennant les contributions du bruit dans ces bandes, ce qui diminuera la variance de l'estimateur de bruit. En outre, ce moyennage permet une forte diminution de la complexité du système.

Les composantes spectrales moyennées $S_{n,i}$ sont adressées à un module 15 de détection d'activité vocale et à un module 16 d'estimation du bruit. Ces deux modules 15, 16 fonctionnent conjointement, en ce sens que des degrés d'activité vocale $\gamma_{n,i}$ mesurés pour les différentes bandes par le module 15 sont utilisés par le module 16 pour estimer l'énergie à long terme du bruit dans les différentes bandes, tandis que ces estimations à long terme $\hat{B}_{n,i}$ sont utilisées par le module 15 pour procéder à un débruitage a priori du signal de parole dans les différentes bandes pour déterminer les degrés d'activité vocale $\gamma_{n,i}$.

Le fonctionnement des modules 15 et 16 peut correspondre aux organigrammes représentés sur les figures 2 et 3.

Aux étapes 17 à 20, le module 15 procède au débruitage a priori du signal de parole dans les différentes bandes i pour la trame de signal n. Ce débruitage a priori est effectué selon un processus classique de soustraction spectrale non linéaire à partir d'estimations du bruit obtenues lors d'une ou plusieurs

trames précédentes. A l'étape 17, le module 15 calcule, avec la résolution des bandes i, la réponse en fréquence $Hp_{n,i}$ du filtre de débruitage a priori, selon la formule :

$$Hp_{n,i} = \frac{S_{n,i} - \alpha'_{n-\tau1,i} \cdot \hat{B}_{n-\tau1,i}}{S_{n-\tau2,i}} \qquad (2)$$

où $\tau1$ et $\tau2$ sont des retards exprimés en nombre de trames ($\tau1 \geq 1$, $\tau2 \geq 0$), et $\alpha'_{n,i}$ est un coefficient de surestimation du bruit dont la détermination sera expliquée plus loin. Le retard $\tau1$ peut être fixe (par exemple $\tau1=1$) ou variable. Il est d'autant plus faible qu'on est confiant dans la détection d'activité vocale.

Aux étapes 18 à 20, les composantes spectrales $\hat{E}p_{n,i}$ sont calculées selon :

$$\hat{E}p_{n,i} = \max\left\{Hp_{n,i} \cdot S_{n,i}\ ,\ \beta p_i \cdot \hat{B}_{n-\tau1,i}\right\} \qquad (3)$$

où βp_i est un coefficient de plancher proche de 0, servant classiquement à éviter que le spectre du signal débruité prenne des valeurs négatives ou trop faibles qui provoqueraient un bruit musical.

Les étapes 17 à 20 consistent donc essentiellement à soustraire du spectre du signal une estimation, majorée par le coefficient $\alpha'_{n-\tau1,i}$, du spectre du bruit estimé a priori.

A l'étape 21, le module 15 calcule l'énergie du signal débruité a priori dans les différentes bandes i pour la trame n : $E_{n,i} = \hat{E}p^2_{n,i}$. Il calcule aussi une moyenne globale $E_{n,0}$ de l'énergie du signal débruité a priori, par une somme des énergies par bande $E_{n,i}$, pondérée par les largeurs de ces bandes. Dans les

notations ci-dessous, l'indice i=0 sera utilisé pour désigner la bande globale du signal.

Aux étapes 22 et 23, le module 15 calcule, pour chaque bande i ($0 \leq i \leq I$), une grandeur $\Delta E_{n,i}$ représentant la variation à court terme de l'énergie du signal débruité dans la bande i, ainsi qu'une valeur à long terme $\bar{E}_{n,i}$ de l'énergie du signal débruité dans la bande i. La grandeur $\Delta E_{n,i}$ peut être calculée par une formule simplifiée de dérivation : $\Delta E_{n,i} = \left| \frac{E_{n-4,i} + E_{n-3,i} - E_{n-1,i} - E_{n,i}}{10} \right|$. Quant à l'énergie à long terme $\bar{E}_{n,i}$, elle peut être calculée à l'aide d'un facteur d'oubli B1 tel que $0<B1<1$, à savoir $\bar{E}_{n,i} = B1 \cdot \bar{E}_{n-1,i} + (1-B1) \cdot E_{n,i}$.

Après avoir calculé les énergies $E_{n,i}$ du signal débruité, ses variations à court terme $\Delta E_{n,i}$ et ses valeurs à long terme $\bar{E}_{n,i}$ de la manière indiquée sur la figure 2, le module 15 calcule, pour chaque bande i ($0 \leq i \leq I$), une valeur ρ_i représentative de l'évolution de l'énergie du signal débruité. Ce calcul est effectué aux étapes 25 à 36 de la figure 3, exécutées pour chaque bande i entre i=0 et i=I. Ce calcul fait appel à un estimateur à long terme de l'enveloppe du bruit ba_i, à un estimateur interne bi_i et à un compteur de trames bruitées b_i.

A l'étape 25, la grandeur $\Delta E_{n,i}$ est comparée à un seuil $\varepsilon 1$. Si le seuil $\varepsilon 1$ n'est pas atteint, le compteur b_i est incrémenté d'une unité à l'étape 26. A l'étape 27, l'estimateur à long terme ba_i est comparé à la valeur de

l'énergie lissée $\bar{E}_{n,i}$. Si $ba_i \geq \bar{E}_{n,i}$, l'estimateur ba_i est pris égal à la valeur lissée $\bar{E}_{n,i}$ à l'étape 28, et le compteur b_i est remis à zéro. La grandeur ρ_i, qui est prise égale au rapport $ba_i/\bar{E}_{n,i}$ (étape 36), est alors égale à 1.

Si l'étape 27 montre que $ba_i < \bar{E}_{n,i}$, le compteur b_i est comparé à une valeur limite bmax à l'étape 29. Si $b_i > bmax$, le signal est considéré comme trop stationnaire pour supporter de l'activité vocale. L'étape 28 précitée, qui revient à considérer que la trame ne comporte que du bruit, est alors exécutée. Si $b_i \leq bmax$ à l'étape 29, l'estimateur interne bi_i est calculé à l'étape 33 selon :

$$bi_i = (1-Bm) \cdot \bar{E}_{n,i} + Bm \cdot ba_i \qquad (4)$$

Dans cette formule, Bm représente un coefficient de mise à jour compris entre 0,90 et 1. Sa valeur diffère selon l'état d'un automate de détection d'activité vocale (étapes 30 à 32). Cet état δ_{n-1} est celui déterminé lors du traitement de la trame précédente. Si l'automate est dans un état de détection de parole ($\delta_{n-1}=2$ à l'étape 30), le coefficient Bm prend une valeur Bmp très proche de 1 pour que l'estimateur du bruit soit très faiblement mis à jour en présence de parole. Dans le cas contraire, le coefficient Bm prend une valeur Bms plus faible, pour permettre une mise à jour plus significative de l'estimateur de bruit en phase de silence. A l'étape 34, l'écart $ba_i - bi_i$ entre l'estimateur à long terme et l'estimateur interne du bruit est comparé à un seuil ε2. Si le seuil ε2 n'est pas atteint, l'estimateur à long terme ba_i est mis à jour avec la valeur de l'estimateur

interne bi_i à l'étape 35. Sinon, l'estimateur à long terme ca_i reste inchangé. On évite ainsi que de brutales variations dues à un signal de parole conduisent à une mise à jour de l'estimateur de bruit.

Après avoir obtenu les grandeurs ρ_i, le module 15 procède aux décisions d'activité vocale à l'étape 37. Le module 15 met d'abord à jour l'état de l'automate de détection selon la grandeur ρ_0 calculée pour l'ensemble de la bande du signal. Le nouvel état δ_n de l'automate dépend de l'état précédent δ_{n-1} et de ρ_0, de la manière représentée sur la figure 4.

Quatre états sont possibles : $\delta=0$ détecte le silence, ou absence de parole ; $\delta=2$ détecte la présence d'une activité vocale ; et les états $\delta=1$ et $\delta=3$ sont des états intermédiaires de montée et de descente. Lorsque l'automate est dans l'état de silence ($\delta_{n-1}=0$), il y reste si ρ_0 ne dépasse pas un premier seuil SE1, et il passe dans l'état de montée dans le cas contraire. Dans l'état de montée ($\delta_{n-1}=1$), il revient dans l'état de silence si ρ_0 est plus petit que le seuil SE1, il passe dans l'état de parole si ρ_0 est plus grand qu'un second seuil SE2 plus grand que le seuil SE1, et il reste dans l'état de montée si $SE1 \leq \rho_0 \leq SE2$. Lorsque l'automate est dans l'état de parole ($\delta_{n-1}=2$), il y reste si ρ_0 dépasse un troisième seuil SE3 plus petit que le seuil SE2, et il passe dans l'état de descente dans le cas contraire. Dans l'état de descente ($\delta_{n-1}=3$), l'automate revient dans l'état de

parole si ρ_0 est plus grand que le seuil SE2, il revient dans l'état de silence si ρ_0 est en deçà d'un quatrième seuil SE4 plus petit que le seuil SE2, et il reste dans l'état de descente si $SE4 \leq \rho_0 \leq SE2$.

A l'étape 37, le module 15 calcule également les degrés d'activité vocale $\gamma_{n,i}$ dans chaque bande $i \geq 1$. Ce degré $\gamma_{n,i}$ est de préférence un paramètre non binaire, c'est-à-dire que la fonction $\gamma_{n,i}=g(\rho_i)$ est une fonction variant continûment entre 0 et 1 en fonction des valeurs prises par la grandeur ρ_i. Cette fonction a par exemple l'allure représentée sur la figure 5.

Le module 16 calcule les estimations du bruit par bande, qui seront utilisées dans le processus de débruitage, en utilisant les valeurs successives des composantes $S_{n,i}$ et des degrés d'activité vocale $\gamma_{n,i}$. Ceci correspond aux étapes 40 à 42 de la figure 3. A l'étape 40, on détermine si l'automate de détection d'activité vocale vient de passer de l'état de montée à l'état de parole. Dans l'affirmative, les deux dernières estimations $\hat{B}_{n-1,i}$ et $\hat{B}_{n-2,i}$ précédemment calculées pour chaque bande $i \geq 1$ sont corrigées conformément à la valeur de l'estimation précédente $\hat{B}_{n-3,i}$. Cette correction est effectuée pour tenir compte du fait que, dans la phase de montée ($\delta=1$), les estimations à long terme de l'énergie du bruit dans le processus de détection d'activité vocale (étapes 30 à 33) ont pu être calculées comme si le signal ne comportait que du bruit (Bm=Bms), de sorte qu'elles risquent d'être entachées d'erreur.

A l'étape 42, le module 16 met à jour les

estimations du bruit par bande selon les formules :

$$\tilde{B}_{n,i} = \lambda_B \cdot \hat{B}_{n-1,i} + (1-\lambda_B) \cdot S_{n,i} \quad (5)$$

$$\hat{B}_{n,i} = \gamma_{n,i} \cdot \hat{B}_{n-1,i} + (1-\gamma_{n,i}) \cdot \tilde{B}_{n,i} \quad (6)$$

où λ_B désigne un facteur d'oubli tel que $0<\lambda_B<1$. La formule (6) met en évidence la prise en compte du degré d'activité vocale non binaire $\gamma_{n,i}$.

Comme indiqué précédemment, les estimations à long terme du bruit $\hat{B}_{n,i}$ font l'objet d'une surestimation, par un module 45 (figure 1), avant de procéder au débruitage par soustraction spectrale non linéaire. Le module 45 calcule le coefficient de surestimation $\alpha'_{n,i}$ précédemment évoqué, ainsi qu'une estimation majorée $\hat{B}'_{n,i}$ qui correspond essentiellement à $\alpha'_{n,i} \cdot \hat{B}_{n,i}$.

L'organisation du module de surestimation 45 est représentée sur la figure 6. L'estimation majorée $\hat{B}'_{n,i}$ est obtenue en combinant l'estimation à long terme $\hat{B}_{n,i}$ et une mesure $\Delta B^{max}_{n,i}$ de la variabilité de la composante du bruit dans la bande i autour de son estimation à long terme. Dans l'exemple considéré, cette combinaison est, pour l'essentiel, une simple somme réalisée par un additionneur 46. Ce pourrait également être une somme pondérée.

Le coefficient de surestimation $\alpha'_{n,i}$ est égal au rapport entre la somme $\hat{B}_{n,i} + \Delta B^{max}_{n,i}$ délivrée par l'additionneur 46 et l'estimation à long terme retardée $\hat{B}_{n-\tau 3,i}$ (diviseur 47), plafonné à une valeur limite α_{max},

par exemple α_{max}=4 (bloc 48). Le retard τ3 sert à corriger le cas échéant, dans les phases de montée (δ=1), la valeur du coefficient de surestimation $\alpha'_{n,i}$, avant que les estimations à long terme aient été corrigées par les étapes 40 et 41 de la figure 3 (par exemple τ3=3).

L'estimation majorée $\hat{B}'_{n,i}$ est finalement prise égale à $\alpha'_{n,i} \cdot \hat{B}_{n-\tau 3,i}$ (multiplieur 49).

La mesure $\Delta B^{max}_{n,i}$ de la variabilité du bruit reflète la variance de l'estimateur de bruit. Elle est obtenue en fonction des valeurs de $S_{n,i}$ et de $\hat{B}_{n,i}$ calculées pour un certain nombre de trames précédentes sur lesquelles le signal de parole ne présente pas d'activité vocale dans la bande i. C'est une fonction des écarts $\left|S_{n-k,i} - \hat{B}_{n-k,i}\right|$ calculés pour un nombre K de trames de silence (n-k≤n). Dans l'exemple représenté, cette fonction est simplement le maximum (bloc 50). Pour chaque trame n, le degré d'activité vocale $\gamma_{n,i}$ est comparé à un seuil (bloc 51) pour décider si l'écart $\left|S_{n,i} - \hat{B}_{n,i}\right|$, calculé en 52-53, doit ou non être chargé dans une file d'attente 54 de K emplacements organisée en mode premier entré-premier sorti (FIFO). Si $\gamma_{n,i}$ ne dépasse pas le seuil (qui peut être égal à 0 si la fonction g() a la forme de la figure 5), la FIFO 54 n'est pas alimentée, tandis qu'elle l'est dans le cas contraire. La valeur maximale contenue dans la FIFO 54 est alors fournie comme mesure de variabilité $\Delta B^{max}_{n,i}$.

La mesure de variabilité $\Delta B^{max}_{n,i}$ peut, en variante,

être obtenue en fonction des valeurs $S_{n,f}$ (et non $S_{n,i}$ et $\hat{B}_{n,i}$. On procède alors de la même manière, sauf que la FIFO 54 contient non pas $\left|S_{n-k,i} - \hat{B}_{n-k,i}\right|$ pour chacune des bandes i, mais plutôt $\max_{f \in [f(i-1), f(i)[} \left|S_{n-k,f} - \hat{B}_{n-k,i}\right|$.

Grâce aux estimations indépendantes des fluctuations à long terme du bruit $\hat{B}_{n,i}$ et de sa variabilité à court terme $\Delta B_{n,i}^{max}$, l'estimateur majoré $\hat{B}'_{n,i}$ procure une excellente robustesse aux bruits musicaux du procédé de débruitage.

Une première phase de la soustraction spectrale est réalisée par le module 55 représenté sur la figure 1. Cette phase fournit, avec la résolution des bandes i ($1 \leq i \leq I$), la réponse en fréquence $H'_{n,i}$ d'un premier filtre de débruitage, en fonction des composantes $S_{n,i}$ et $\hat{B}_{n,i}$ et des coefficients de surestimation $\alpha'_{n,i}$. Ce calcul peut être effectué pour chaque bande i selon la formule :

$$H'_{n,i} = \frac{\max\left\{S_{n,i} - \alpha'_{n,i} \cdot \hat{B}_{n,i} \, , \, \beta'_i \cdot \hat{B}_{n,i}\right\}}{S_{n-\tau 4,i}} \qquad (7)$$

où $\tau 4$ est un retard entier déterminé tel que $\tau 4 \geq 0$ (par exemple $\tau 4 = 0$). Dans l'expression (7), le coefficient β'_i représente, comme le coefficient βp_i de la formule (3), un plancher servant classiquement à éviter les valeurs négatives ou trop faibles du signal débruité.

De façon connue (EP-A-0 534 837), le coefficient de surestimation $\alpha'_{n,i}$ pourrait être remplacé dans la

formule (7) par un autre coefficient égal à une fonction de $\alpha'_{n,i}$ et d'une estimation du rapport signal-sur-bruit (par exemple $S_{n,i}/\hat{B}_{n,i}$), cette fonction étant décroissante selon la valeur estimée du rapport signal-sur-bruit. Cette fonction est alors égale à $\alpha'_{n,i}$ pour les valeurs les plus faibles du rapport signal-sur-bruit. En effet, lorsque le signal est très bruité, il n'est a priori pas utile de diminuer le facteur de surestimation. Avantageusement, cette fonction décroît vers zéro pour les valeurs les plus élevées du rapport signal/bruit. Ceci permet de protéger les zones les plus énergétiques du spectre, où le signal de parole est le plus significatif, la quantité soustraite du signal tendant alors vers zéro.

Cette stratégie peut être affinée en l'appliquant de manière sélective aux harmoniques de la fréquence tonale (« pitch ») du signal de parole lorsque celui-ci présente une activité vocale.

Ainsi, dans la réalisation représentée sur la figure 1, une seconde phase de débruitage est réalisée par un module 56 de protection des harmoniques. Ce module calcule, avec la résolution de la transformée de Fourier, la réponse en fréquence $H^2_{n,f}$ d'un second filtre de débruitage en fonction des paramètres $H^1_{n,i}$, $\alpha'_{n,i}$, $\hat{B}_{n,i}$, δ_n, $S_{n,i}$ et de la fréquence tonale $f_p=F_e/T_p$ calculée en dehors des phases de silence par un module d'analyse harmonique 57. En phase de silence ($\delta_n=0$), le module 56 n'est pas en service, c'est-à-dire que $H^2_{n,f} = H^1_{n,i}$ pour chaque fréquence f d'une bande i. Le module 57 peut appliquer toute méthode connue d'analyse du signal de

parole de la trame pour déterminer la période T_p, exprimée comme un nombre entier ou fractionnaire d'échantillons, par exemple une méthode de prédiction linéaire.

La protection apportée par le module 56 peut consister à effectuer, pour chaque fréquence f appartenant à une bande i :

$$\begin{cases} H^2_{n,f} = 1 \quad \text{si} \quad \begin{cases} S_{n,i} - \alpha'_{n,i} \cdot \hat{B}_{n,i} > \beta^2_i \cdot \hat{B}_{n,i} & (8) \\ \text{et } \exists \eta \text{ entier } / \left| f - \eta . f_p \right| \leq \Delta f / 2 & (9) \end{cases} \\ H^2_{n,f} = H^1_{n,f} \quad \text{sinon} \end{cases}$$

$\Delta f = F_e/N$ représente la résolution spectrale de la transformée de Fourier. Lorsque $H^2_{n,f}=1$, la quantité soustraite de la composante $S_{n,f}$ sera nulle. Dans ce calcul, les coefficients de plancher β^2_i (par exemple $\beta^2_i = \beta^1_i$) expriment le fait que certaines harmoniques de la fréquence tonale f_p peuvent être masquées par du bruit, de sorte qu'il n'est pas utile de les protéger.

Cette stratégie de protection est de préférence appliquée pour chacune des fréquences les plus proches des harmoniques de f_p, c'est-à-dire pour η entier quelconque.

Si on désigne par δf_p la résolution fréquentielle avec laquelle le module d'analyse 57 produit la fréquence tonale estimée f_p, c'est-à-dire que la fréquence tonale réelle est comprise entre $f_p - \delta f_p/2$ et $f_p + \delta f_p/2$, alors l'écart entre la η-ième harmonique de la fréquence tonale réelle est son estimation $\eta \times f_p$ (condition (9)) peut aller jusqu'à $\pm \eta \times \delta f_p/2$. Pour les valeurs élevées de η, cet écart peut être supérieur à la demi-résolution spectrale $\Delta f/2$ de

la transformée de Fourier. Pour tenir compte de cette incertitude et garantir la bonne protection des harmoniques de la fréquence tonale réelle, on peut protéger chacune des fréquences de l'intervalle $\left[\eta \times f_p - \eta \times \delta f_p/2 \ , \eta \times f_p + \eta \times \delta f_p/2\right]$, c'est-à-dire remplacer la condition (9) ci-dessus par :

$$\exists \eta \text{ entier } / \left| f - \eta . f_p \right| \leq \left(\eta . \delta f_p + \Delta f \right)/2 \qquad (9')$$

Cette façon de procéder (condition (9')) présente un intérêt particulier lorsque les valeurs de η peuvent être grandes, notamment dans le cas où le procédé est utilisé dans un système à bande élargie.

Pour chaque fréquence protégée, la réponse en fréquence corrigée $H^2_{n,f}$ peut être égale à 1 comme indiqué ci-dessus, ce qui correspond à la soustraction d'une quantité nulle dans le cadre de la soustraction spectrale, c'est-à-dire à une protection complète de la fréquence en question. Plus généralement, cette réponse en fréquence corrigée $H^2_{n,f}$ pourrait être prise égale à une valeur comprise entre 1 et $H^1_{n,f}$ selon le degré de protection souhaité, ce qui correspond à la soustraction d'une quantité inférieure à celle qui serait soustraite si la fréquence en question n'était pas protégée.

Les composantes spectrales $S^2_{n,f}$ d'un signal débruité sont calculées par un multiplieur 58 :

$$S^2_{n,f} = H^2_{n,f} . S_{n,f} \qquad (10)$$

Ce signal $S^2_{n,f}$ est fourni à un module 60 qui calcule, pour chaque trame n, une courbe de masquage en appliquant un modèle psychoacoustique de perception

auditive par l'oreille humaine.

Le phénomène de masquage est un principe connu du fonctionnement de l'oreille humaine. Lorsque deux fréquences sont entendues simultanément, il est possible que l'une des deux ne soit plus audible. On dit alors qu'elle est masquée.

Il existe différentes méthodes pour calculer des courbes de masquage. On peut par exemple utiliser celle développée par J.D. Johnston («Transform Coding of Audio Signals Using Perceptual Noise Criteria », IEEE Journal on Selected Area in Communications, Vol. 6, No. 2, février 1988). Dans cette méthode, on travaille dans l'échelle fréquentielle des barks. La courbe de masquage est vue comme la convolution de la fonction d'étalement spectral de la membrane basilaire dans le domaine bark avec le signal excitateur, constitué dans la présente application par le signal $S^2_{n,f}$. La fonction d'étalement spectral peut être modélisée de la manière représentée sur la figure 7. Pour chaque bande de bark, on calcule la contribution des bandes inférieures et supérieures convoluées par la fonction d'étalement de la membrane basilaire :

$$C_{n,q} = \sum_{q'=0}^{q-1} \frac{S^2_{n,q'}}{\left(10^{10/10}\right)^{(q-q')}} + \sum_{q'=q+1}^{Q} \frac{S^2_{n,q'}}{\left(10^{25/10}\right)^{(q'-q)}} \qquad (11)$$

où les indices q et q' désignent les bandes de bark ($0 \le q, q' \le Q$), et $S^2_{n,q'}$ représente la moyenne des composantes $S^2_{n,f}$ du signal excitateur débruité pour les fréquences discrètes f appartenant à la bande de bark q'.

Le seuil de masquage $M_{n,q}$ est obtenu par le module pour chaque bande de bark q, selon la formule :

$$M_{n,q} = C_{n,q}/R_q \qquad (12)$$

où R_q dépend du caractère plus ou moins voisé du signal. De façon connue, une forme possible de R_q est :

$$10.\log_{10}(R_q) = (A+q).\chi + B.(1-\chi) \qquad (13)$$

avec A=14,5 et B=5,5. χ désigne un degré de voisement du signal de parole, variant entre zéro (pas de voisement) et 1 (signal fortement voisé). Le paramètre χ peut être de la forme connue :

$$\chi = \min\left\{\frac{SFM}{SFM_{max}}, 1\right\} \qquad (12)$$

où SFM représente, en décibels, le rapport entre la moyenne arithmétique et la moyenne géométrique de l'énergie des bandes de bark, et SFM_{max}=-60 dB.

Le système de débruitage comporte encore un module 62 qui corrige la réponse en fréquence du filtre de débruitage, en fonction de la courbe de masquage $M_{n,q}$ calculée par le module 60 et des estimations majorées $\hat{B}'_{n,i}$ calculées par le module 45. Le module 62 décide du niveau de débruitage qui doit réellement être atteint.

En comparant l'enveloppe de l'estimation majorée du bruit avec l'enveloppe formée par les seuils de masquage $M_{n,q}$, on décide de ne débruiter le signal que dans la mesure où l'estimation majorée $\hat{B}'_{n,i}$ dépasse la courbe de masquage. Ceci évite de supprimer inutilement du bruit masqué par de la parole.

La nouvelle réponse $H^3_{n,f}$, pour une fréquence f appartenant à la bande i définie par le module 12 et à la bande de bark q, dépend ainsi de l'écart relatif entre

l'estimation majorée $\hat{B}'_{n,i}$ de la composante spectrale correspondante du bruit et la courbe de masquage $M_{n,q}$, de la manière suivante :

$$H^3_{n,f} = 1 - \left(1 - H^2_{n,f}\right) \cdot \max\left\{\frac{\hat{B}'_{n,i} - M_{n,q}}{\hat{B}'_{n,i}}, 0\right\} \qquad (14)$$

En d'autres termes, la quantité soustraite d'une composante spectrale $S_{n,f}$, dans le processus de soustraction spectrale ayant la réponse fréquentielle $H^3_{n,f}$, est sensiblement égale au minimum entre d'une part la quantité soustraite de cette composante spectrale dans le processus de soustraction spectrale ayant la réponse fréquentielle $H^2_{n,f}$, et d'autre part la fraction de l'estimation majorée $\hat{B}'_{n,i}$ de la composante spectrale correspondante du bruit qui, le cas échéant, dépasse la courbe de masquage $M_{n,q}$.

La figure 8 illustre le principe de la correction appliquée par le module 62. Elle montre schématiquement un exemple de courbe de masquage $M_{n,q}$ calculée sur la base des composantes spectrales $S^2_{n,f}$ du signal débruité, ainsi que l'estimation majorée $\hat{B}'_{n,i}$ du spectre du bruit. La quantité finalement soustraite des composantes $S_{n,f}$ sera celle représentée par les zones hachurées, c'est-à-dire limitée à la fraction de l'estimation majorée $\hat{B}'_{n,i}$ des composantes spectrales du bruit qui dépasse la courbe de masquage.

Cette soustraction est effectuée en multipliant la réponse fréquentielle $H^3_{n,f}$ du filtre de débruitage par

les composantes spectrales $S_{n,\xi}$ du signal de parole (multiplieur 64). Un module 65 reconstruit alors le signal débruité dans le domaine temporel, en opérant la transformée de Fourier rapide inverse (TFRI) inverse des échantillons de fréquence $S^3_{n,\xi}$ délivrés par le multiplieur 64. Pour chaque trame, seuls les N/2=128 premiers échantillons du signal produit par le module 65 sont délivrés comme signal débruité final s^3, après reconstruction par addition-recouvrement avec les N/2=128 derniers échantillons de la trame précédente (module 66).

REVENDICATIONS

1. Procédé de détection d'activité vocale dans un signal de parole numérique (s) traité par trames successives, dans lequel on soumet le signal de parole à un débruitage en tenant compte d'estimations du bruit compris dans le signal, mises à jour pour chaque trame d'une manière dépendante d'au moins un degré d'activité vocale ($\gamma_{n,i}$) déterminé pour ladite trame, caractérisé en ce qu'on procède à un débruitage a priori du signal de parole de chaque trame sur la base d'estimations du bruit ($\alpha'_{n-1,i} \cdot \hat{B}_{n-1,i}$) obtenues lors du traitement d'au moins une trame précédente, et on analyse les variations d'énergie du signal débruité a priori ($\hat{E}p_{n,i}$) pour détecter le degré d'activité vocale de ladite trame.

2. Procédé selon la revendication 1, dans lequel le degré d'activité vocale ($\gamma_{n,i}$) est un paramètre non binaire.

3. Procédé selon la revendication 2, dans lequel le degré d'activité vocale ($\gamma_{n,i}$) est une fonction, variant continûment entre 0 et 1.

4. Procédé selon l'une quelconque des revendications précédentes, dans lequel les estimations du bruit sont obtenues dans différentes bandes fréquentielles du signal, le débruitage a priori est effectué bande par bande, et il est déterminé un degré d'activité vocale ($\gamma_{n,i}$) pour chaque bande.

5. Procédé selon l'une quelconque des revendications précédentes, dans lequel on obtient une estimation du bruit $\hat{B}_{n,i}$ pour la trame n dans une bande de fréquences i sous la forme :

$$\hat{B}_{n,i} = \gamma_{n,i} \cdot \hat{B}_{n-1,i} + (1-\gamma_{n,i}) \cdot \tilde{B}_{n,i}$$

avec $\tilde{B}_{n,i} = \lambda_B \cdot \hat{B}_{n-1,i} + (1-\lambda_B) \cdot S_{n,i}$

où λ_B est un facteur d'oubli compris entre 0 et 1, $\gamma_{n,i}$ est le degré d'activité vocale déterminé pour la trame n dans la bande de fréquences i, et $S_{n,i}$ est une moyenne de l'amplitude du spectre du signal de parole de la trame n sur la bande i.

6. Procédé selon la revendication 5, dans lequel le signal débruité a priori $\hat{E}p_{n,i}$ relativement à une trame n et à une bande de fréquences i est de la forme :

$$\hat{E}p_{n,i} = \max\left\{Hp_{n,i} \cdot S_{n,i} \, , \, \beta p_i \cdot \hat{B}_{n-\tau 1,i}\right\}$$

où $Hp_{n,i} = \dfrac{S_{n,i} - \alpha'_{n-\tau 1,i} \cdot \hat{B}_{n-\tau 1,i}}{S_{n-\tau 2,i}}$, $\tau 1$ est un entier au moins égal à 1, $\tau 2$ est un entier au moins égal à 0, $\alpha'_{n-\tau 1,i}$ est un coefficient de surestimation déterminé pour la trame $n-\tau 1$ et la bande i, et βp_i est un coefficient positif.

7. Procédé selon l'une quelconque des revendications précédentes, dans lequel on calcule une estimation à long terme ($\bar{E}_{n,i}$) de l'énergie du signal débruité a priori ($\hat{E}p_{n,i}$), et on compare cette estimation à long terme à une estimation instantanée (ba) de cette énergie, calculée sur la trame en cours, pour obtenir le degré d'activité vocale ($\gamma_{n,i}$) de ladite trame.

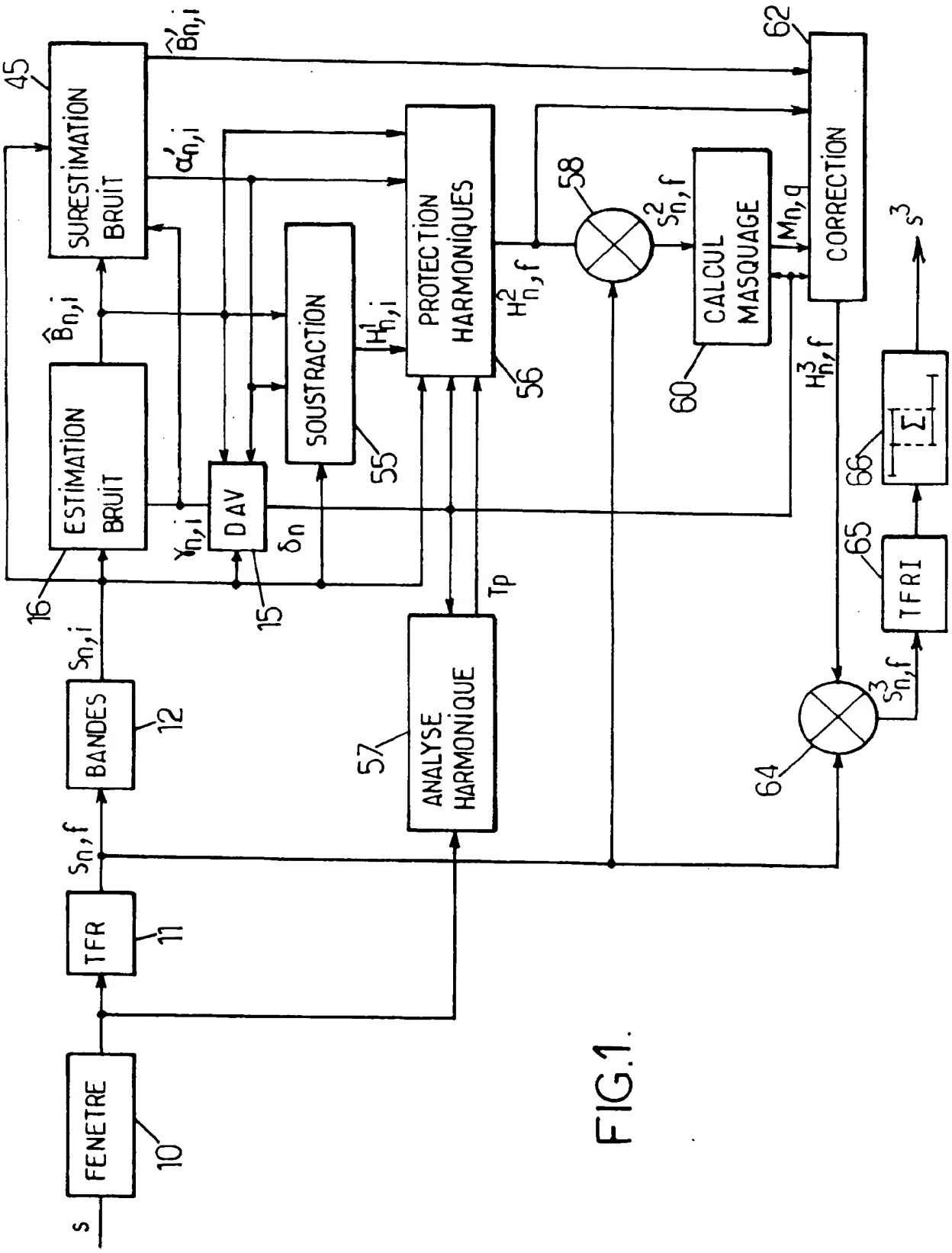


FIG.1.

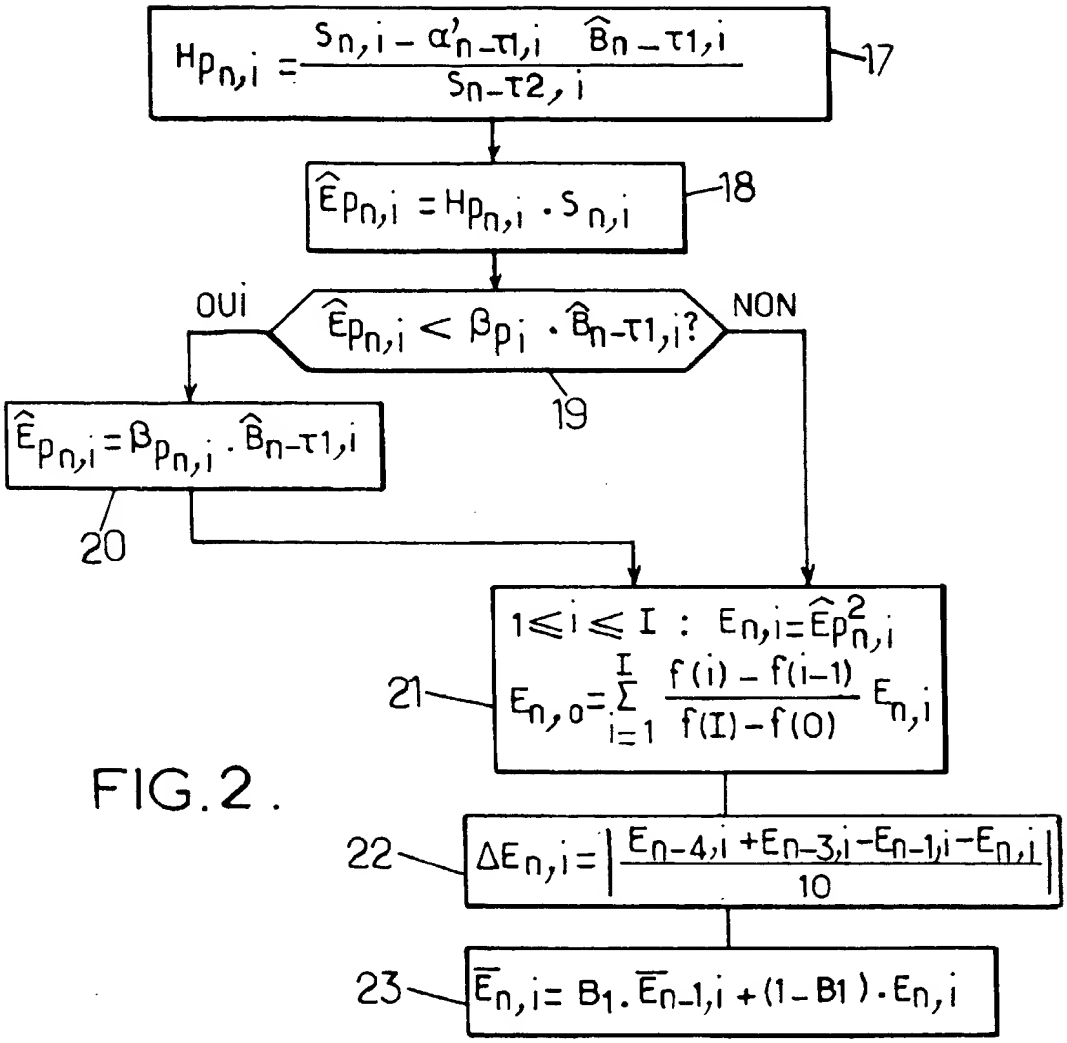


FIG.2.

FIG.4.

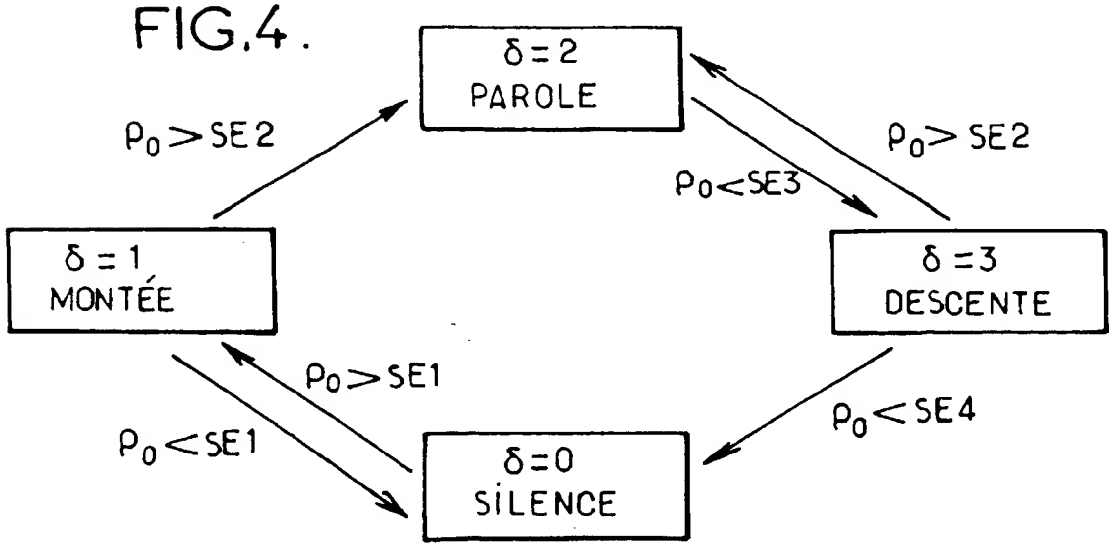
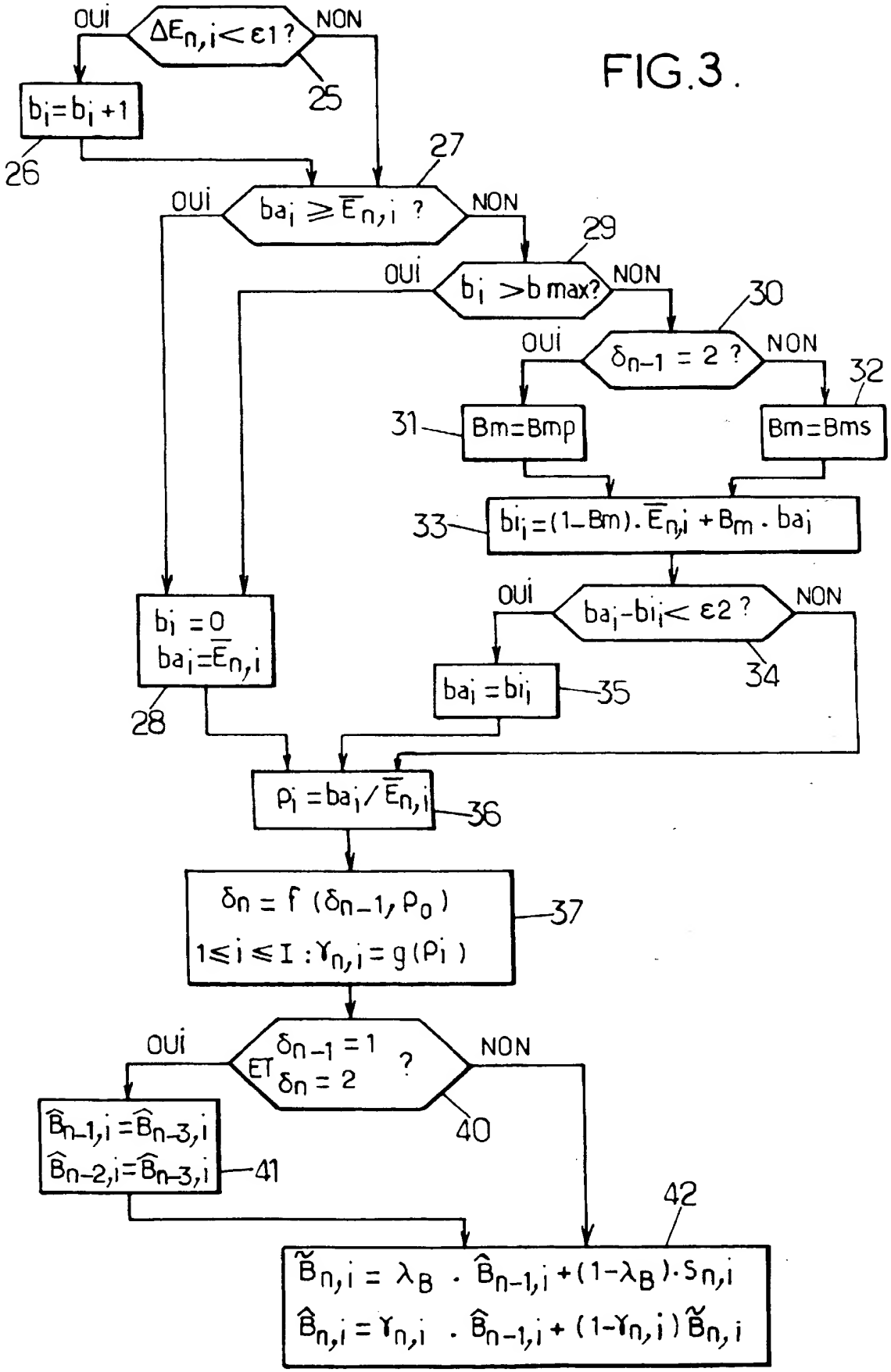


FIG.3.



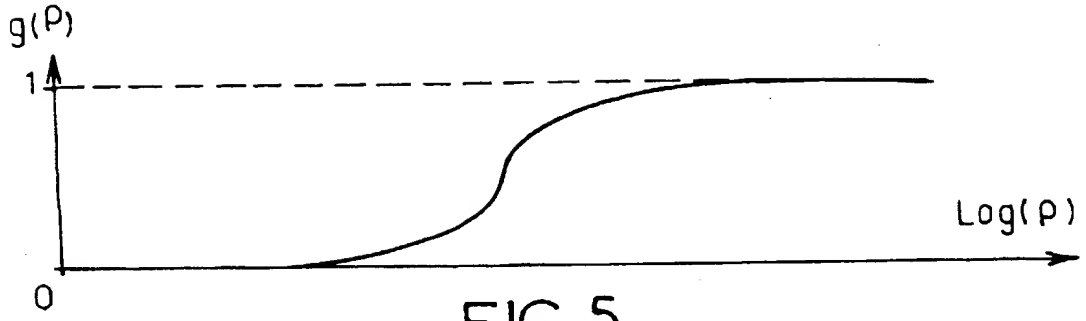


FIG.5.

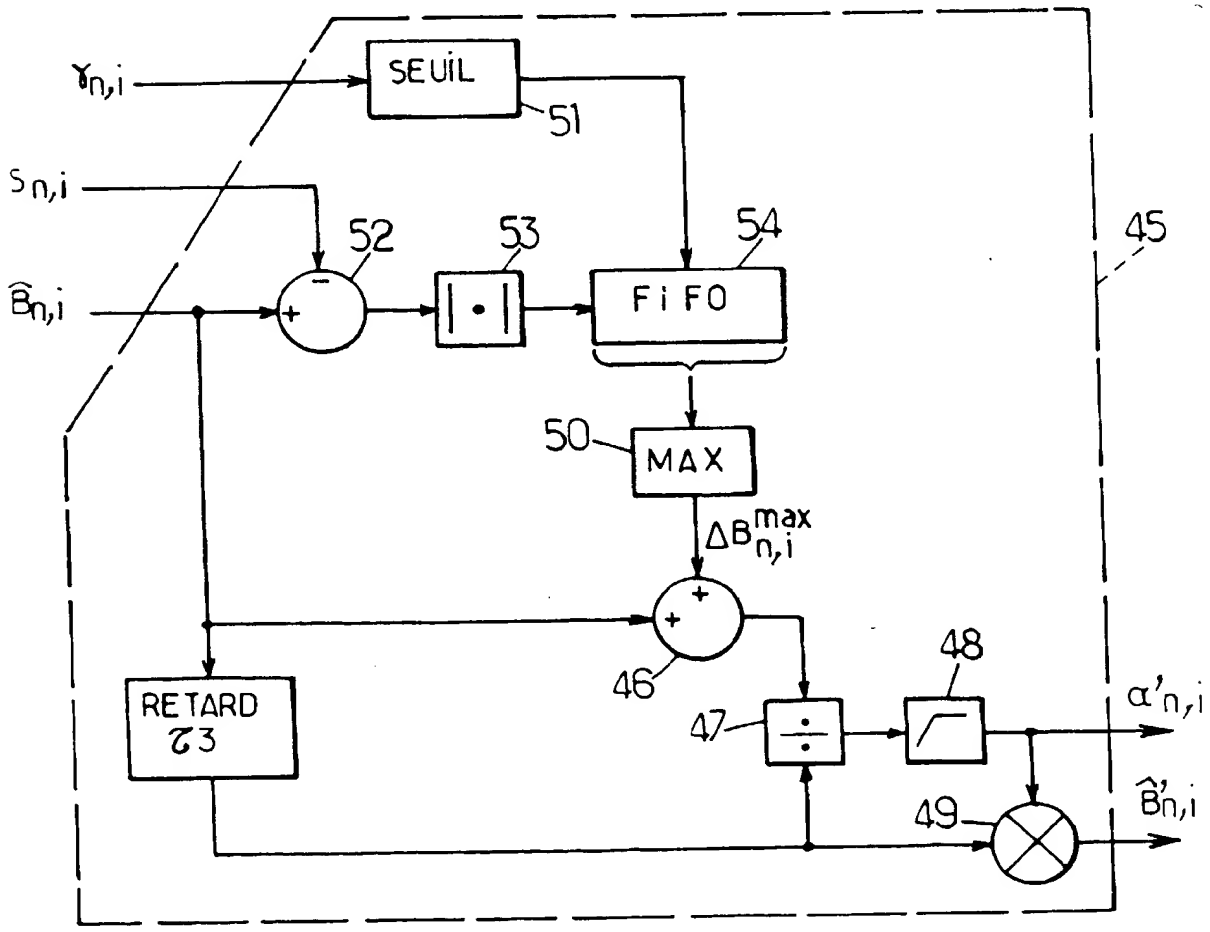


FIG.6.

FIG. 7.

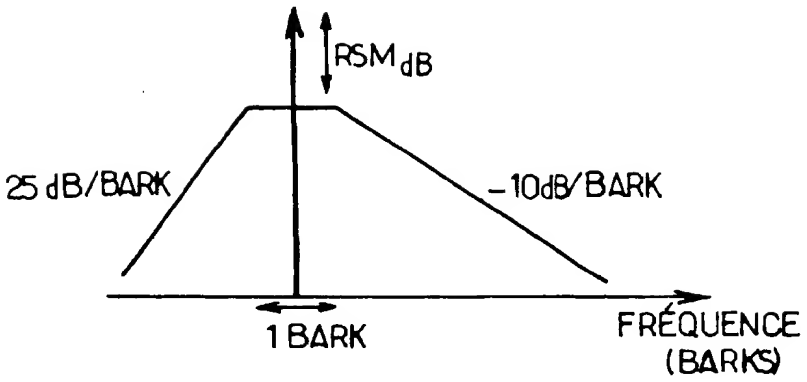
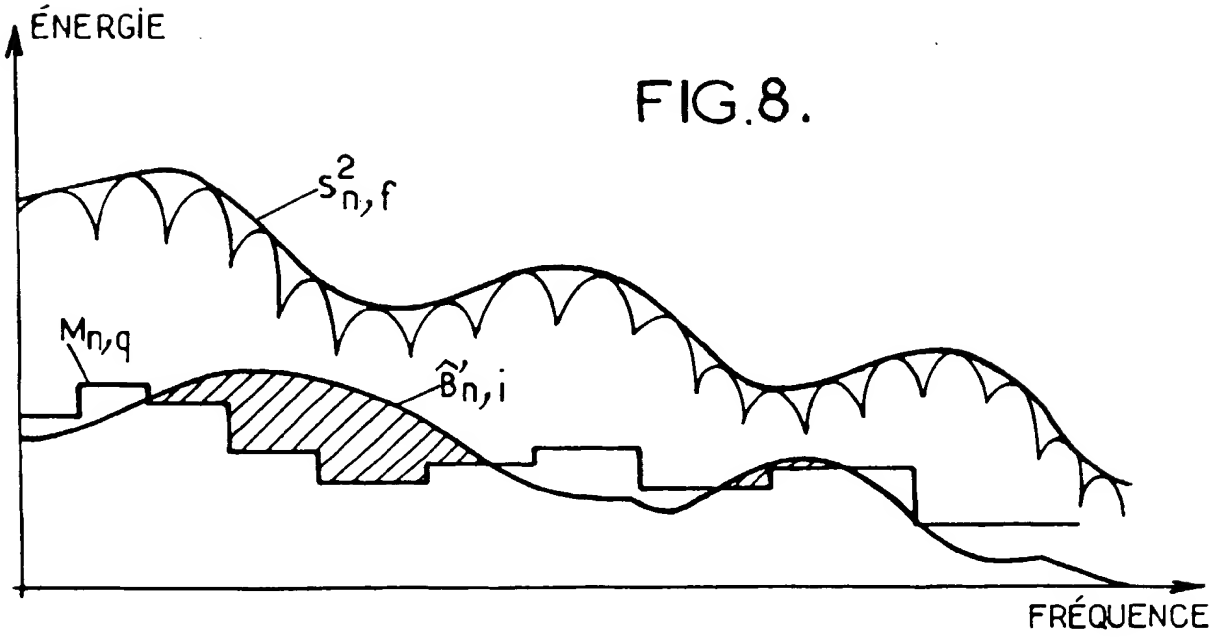


FIG. 8.



REPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
d la
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE PRELIMINAIRE

établi sur la bas des d mi`res revendications
déposées avant le commenc ment de la recherche

2768544
N° d'enregistrement national

FA 549901
FR 9711640

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	
A	US 5 659 622 A (ASHLEY) 19 août 1997 * colonne 3, ligne 12 - ligne 53 * ---	1
A	PATENT ABSTRACTS OF JAPAN vol. 095, no. 006, 31 juillet 1995 & JP 07 074709 A (SONY), 17 mars 1995, * abrégé * -& US 5 732 390 A (KATAYANAGI ET AL.) 24 mars 1998 * colonne 2, ligne 17 - ligne 56 * * colonne 6 - colonne 7, ligne 28 * ---	1
A	DE 40 12 349 A (RICOH) 25 octobre 1990 * colonne 4 - colonne 5 * -----	1

DOMAINES TECHNIQUES RECHERCHES (Int.CL.6)
G10L

1

Date d'achèvement de la recherche	Examinateur
18 juin 1998	Lange, J

CATEGORIE DES DOCUMENTS CITES

X : particulièrement pertinent à lui seul
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie
A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général
O : divulgation non-écrite
P : document intercalaire

T : théorie ou principe à la base de l'invention
E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.
D : cité dans la demande
L : cité pour d'autres raisons

& : membre de la même famille, document correspondant

EPO FORM 1503 03.82 (P04C13)